

# Coarse-grained sensitivity for multiscale data assimilation

Nozomi Sugiura\*

*Japan Agency for Marine-Earth Science and Technology*

We show that the effective average action and its gradient are useful for solving multiscale data assimilation problems. We also present a procedure for numerically evaluating the gradient of the effective average action, and demonstrate that the variational problem for slow degrees of freedom can be solved properly using the “effective gradient.”

PACS numbers: 05.45.Tp 05.10.Cc 91.10.Vr

arXiv:1510.09026v5 [physics.data-an] 25 Mar 2016

---

\* nsugiura@jamstec.go.jp

## I. INTRODUCTION

A key problem in the assimilation of data for nonlinear multiscale systems concerns the optimization of the slow degrees of freedom after the fast degrees of freedom have been properly averaged [1]. This is also the case with data assimilation for coupled atmosphere–ocean systems [2]. From a statistical point of view, this amounts to performing some integration with respect to similar realizations of control variables, and packing them together into an “effective” cost function (action) [3] (see below for the definition). Geometrically, the rough surface of the original cost function can be smoothed according to a coarse-grained averaging procedure.

To see this, we first review how the concept of the effective action is relevant to data assimilation [4]. Data assimilation concerns the following statistical problem: given the observation  $y \in \mathbb{R}^p$ , the prior probability  $P(\chi)$  of the control variable  $\chi \in \mathbb{R}^M$ , and the likelihood  $P(y|\chi)$  of the observation, the conditional expectation of any physical quantity  $G(\chi)$  is calculated through the integral:

$$E[G(\chi)|y] = \frac{\int d\chi G(\chi) P(\chi) P(y|\chi)}{\int d\chi P(\chi) P(y|\chi)} = \frac{\int d\chi G(\chi) e^{-S[\chi]}}{\int d\chi e^{-S[\chi]}}, \quad (1)$$

where  $S[\chi]$  is called the action, or cost function, and

$$\int d\chi = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} d\chi_1 d\chi_2 \cdots d\chi_M, \quad (2)$$

denotes the multiple integral over all possible combinations (paths) of  $\chi$ , also called the path integral. Although the control variable  $\chi = \chi(x, t)$  can generally be a field defined in some space-time  $(x, t)$ , we confine ourselves to the case of a discrete space-time with  $M$  cells, that is,  $\chi \in \mathbb{R}^M$ . Note that Eq. (1) includes the posterior probability  $P(\chi|y) = P(\chi)P(y|\chi)/P(y)$  as a special case with the delta functional  $G(\chi') = \delta(\chi' - \chi)$ .

If the posterior  $P(\chi|y)$  is highly concentrated around the most probable state  $\hat{\chi}$ , which means  $P(\chi|y) \simeq \delta(\chi - \hat{\chi})$ , Eq. (1) can be approximated as:

$$E[G(\chi)|y] \simeq G(\hat{\chi}). \quad (3)$$

In this situation, it is important to find the control variable that minimizes the cost function  $S[\chi]$ . 4D-Var efficiently determines one of the stationary points satisfying  $\delta S[\chi]/\delta\chi = 0$ . However, it does not necessarily give the desired global minimum, as the cost function may have multiple minima. With regard to the shape of the cost function, we could fail to see the forest for the trees.

To deal with more general posterior probabilities, and to calculate the conditional expectation more robustly, we present an effective alternative to the cost function. We introduce an external source term  $-J^T \chi$  ( $J \in \mathbb{R}^M$  is an external field) to the action in the normalization factor  $\int d\chi \exp(-S[\chi])$  of Eq. (1). This leads to the following definition of the partition function:

$$Z[J] = \int d\chi e^{-S[\chi] + J^T \chi}, \quad (4)$$

which encodes all the information about the conditional expectation as follows:

$$E[G(\chi)|y] = \frac{1}{Z(0)} G\left(\frac{\delta}{\delta J}\right) Z[J] \Big|_{J=0}, \quad (5)$$

where  $G(\delta/\delta J)$  should be interpreted as an operator in which the argument  $\chi$  of the algebraic expression  $G(\chi)$  is replaced with the differential operator. The logarithm of the partition function:

$$W[J] = \log Z[J] \quad (6)$$

is also useful, because it contains all the information about the cumulants. For example,

$$\frac{\delta W}{\delta J}[0] = E[\chi^T | y], \quad (7)$$

$$\frac{\delta^2 W}{\delta J^2}[0] = E\left[(\chi - E[\chi|y])(\chi - E[\chi|y])^T | y\right]. \quad (8)$$

That is to say, we can extract information about the expected value if we perturb the external field  $J$  and observe how the normalization factor  $Z[J]$ , or  $W[J]$ , changes.

To estimate the expected value, we can construct a functional called the effective action [3, 4], whose independent variable is the expected value  $\phi$  in the presence of the external field  $J$ , through the Legendre transformation:

$$\Gamma[\phi] \equiv \sup_J \{-W[J] + J^T \phi\}. \quad (9)$$

Taking the supremum while  $\phi$  remains fixed in Eq. (9), we obtain

$$\phi^T = \frac{\delta W}{\delta J}[J]. \quad (10)$$

Using this and taking the derivative with respect to  $\phi$  gives

$$\frac{\delta \Gamma}{\delta \phi}[\phi] = -\frac{\delta W}{\delta J}[J] \frac{\delta J}{\delta \phi} + \frac{\delta W}{\delta J}[J] \frac{\delta J}{\delta \phi} + J^T = J^T. \quad (11)$$

There are at least two advantages to this transformation. First, it results in a convex function because we take the Legendre transform of a convex function  $W[J]$  [5]. Second, it elicits a symmetric relation between  $W[J]$  and  $\Gamma[\phi]$ :

$$\phi^T = \frac{\delta W}{\delta J}[J], \quad J^T = \frac{\delta \Gamma}{\delta \phi}[\phi]. \quad (12)$$

This implies that its unique stationary point, which satisfies  $\delta \Gamma[\phi]/\delta \phi = 0$ , identifies the expected value  $\phi = \delta W[0]/\delta J = E[\chi|y]$ . In other words, we can find the conditional expectation by finding the stationary point of the effective action. Eq. (12) also implies

$$\frac{\delta^2 \Gamma}{\delta \phi^2}[\phi] = \frac{\delta J}{\delta \phi} = \left( \frac{\delta \phi}{\delta J} \right)^{-1} = \left( \frac{\delta^2 W}{\delta J^2}[J] \right)^{-1}. \quad (13)$$

Comparing this with (8), we see that the stationary point of the effective action also provides more cumulant information.

Since  $\Gamma[\phi]$  in Eq. (9) should be regarded as a function of  $\phi$  alone, we eliminate  $J$  using Eq. (12) to obtain

$$\Gamma[\phi] = -W \left[ \frac{\delta \Gamma}{\delta \phi}[\phi] \right] + \frac{\delta \Gamma}{\delta \phi}[\phi] \phi \quad (14)$$

$$= -\log \left\{ \int d\chi e^{-S[\chi] + \frac{\delta \Gamma}{\delta \phi}[\phi] \chi} \right\} + \frac{\delta \Gamma}{\delta \phi}[\phi] \phi \quad (15)$$

$$= -\log \left\{ \int d\chi e^{-S[\chi] + \frac{\delta \Gamma}{\delta \phi}[\phi] (\chi - \phi)} \right\}. \quad (16)$$

This suggests a means of calculating the effective action. However, it requires a recursive procedure that includes integrations over all possible combinations of control variables  $\chi$ , which appears to be intractable.

To compute the effective action, we may evaluate this integral stepwise, using methods developed in renormalization group theory [6]. A relevant concept that we shall explore later is the effective average action  $\Gamma_k[\phi]$  proposed by Wetterich [7], which constitutes a one-parameter family of functionals interpolating between the action  $S[\phi]$  and the effective action  $\Gamma[\phi]$ .

The aim of this paper is to propose a possible framework that will help solve the multiscale data assimilation problem by replacing the cost function with the effective average action. We also propose a novel method for evaluating the gradient of the effective average action numerically. In principle, this enables us to solve a broader range of data assimilation problems by seeking the stationary point of the effective average action using its gradient, which is referred to as the effective gradient or the coarse-grained sensitivity.

The concept of the effective average action is explained in section II A, and the meaning of its stationary point is clarified in section II B. Section III describes a novel procedure for calculating the gradient of the effective action. Sections IV and V illustrate some applications of the method to data assimilation or sensitivity studies.

## II. EFFECTIVE AVERAGE ACTION

When dealing with a multiscale system, it is often difficult to define the sensitivity with respect to the control variable, because fast degrees of freedom may have many statistical paths related to the sensitivity that cannot be expressed in a deterministic manner. In other words, we cannot always use the sensitivity to choose the optimal realization of a fluctuation from among a large ensemble of fast fluctuations in the control space. This motivates the definition of a macroscopic field in which fast degrees of freedom are treated as averaged quantities. A suitable tool for this purpose is the effective average action [7].

### A. Definition

The procedure for the effective action reviewed in the Introduction,  $S[\chi] \rightarrow W[J] \rightarrow \Gamma[\phi]$ , can also be applied to the derivation of the effective average action. This introduces some filtering terms,  $\Delta S_k[\chi]$  to Eq. (4) and  $-\Delta S_k[\phi]$  to Eq. (9), which have the effect of selectively integrating out the fast degrees of freedom in the control space to enable the dynamics of slower variables to be investigated.

We start by defining an infra-red filter[8],

$$\Delta S_k[\chi] \equiv \frac{1}{2} \chi^T R_k \chi, \quad (17)$$

where  $\chi \in \mathbb{R}^M$  is the control variable in a discrete space-time with  $M$  cells, and  $R_k \in \mathbb{R}^M \times \mathbb{R}^M$  is a discrete low-pass filter.

To derive concrete expressions for  $R_k$ , let us consider a simple case with a cyclic control variable  $\chi$  in a 1-dimensional discrete domain  $l = 1, 2, \dots, M$ . We define the discrete Fourier transform  $\widehat{\chi}$  of  $\chi$  and its inverse as:

$$\widehat{\chi}_j = \frac{1}{\sqrt{M}} \sum_{l=1}^M \chi_l e^{-\frac{2\pi j l}{M} i}, \quad |j| < [M/2], \quad (18)$$

$$\chi_l = \frac{1}{\sqrt{M}} \sum_{j=-[M/2]}^{[M/2]} \widehat{\chi}_j e^{\frac{2\pi j l}{M} i}, \quad l = 1, 2, \dots, M, \quad (19)$$

where  $M$  is odd for simplicity,  $[\cdot]$  denotes the roundoff, and  $i$  is the imaginary unit. If we assume the infra-red filter is represented by a cutoff of high-wavenumber modes, then:

$$\Delta S_k[\widehat{\chi}] = \frac{1}{2} \sum_j \widehat{\chi}_{-j} \widehat{R}_k(j) \widehat{\chi}_j, \quad (20)$$

$$\widehat{R}_k(j) = \begin{cases} k^2 \left(1 - \frac{j^2}{j_k^2}\right) & \text{if } |j| < j_k \\ 0 & \text{otherwise} \end{cases}, \quad (21)$$

where  $j, j_k \in \mathbb{Z}$ , and  $j_k$  is the cutoff level. The filtering term can then be written as

$$\begin{aligned} \Delta S_k[\widehat{\chi}] &= \frac{1}{2} \sum_{|j| < j_k} \widehat{\chi}_{-j} k^2 \left(1 - \frac{j^2}{j_k^2}\right) \widehat{\chi}_j, \\ &= \frac{1}{2} \sum_{l, l'=1}^M \chi_l \left\{ \sum_{|j| < j_k} \frac{k^2}{M} \left(1 - \frac{j^2}{j_k^2}\right) e^{\frac{-2\pi j(l-l')}{M} i} \right\} \chi_{l'} \end{aligned} \quad (22)$$

$$= \frac{1}{2} \sum_{l, l'=1}^M \chi_l \left\{ \sum_{|j| < j_k} \frac{k^2}{M} \left(1 - \frac{j^2}{j_k^2}\right) \cos \left[ \frac{2\pi j(l-l')}{M} \right] \right\} \chi_{l'}. \quad (23)$$

The expression in the curly brackets gives a matrix representation of  $R_k$  in position space. If  $M$  is even, we can replace Eq. (21) by

$$\widehat{R}_k \left( j + \frac{1}{2} \right) = \begin{cases} k^2 \left(1 - \frac{(j+\frac{1}{2})^2}{j_k^2}\right) & \text{if } |j + \frac{1}{2}| < j_k \\ 0 & \text{otherwise} \end{cases} \quad (24)$$

where  $j, j_k \in \mathbb{Z}$ .

To further simplify the filter, we can also use  $R_k = k^2$  with  $j_k \rightarrow \infty$ , which yields

$$\Delta S_k[\chi] = \frac{k^2}{2} \sum_{l=1}^M \chi_l^2. \quad (25)$$

As we will see later, the filter should have the following properties:

$$k \rightarrow \infty \Rightarrow R_k \rightarrow \infty, \quad (26)$$

$$k \rightarrow 0 \Rightarrow R_k \rightarrow 0. \quad (27)$$

With the filtering term, the partition function  $Z_k$  and its logarithm for the high-wavenumber modes can be defined as:

$$Z_k[J] = \int d\chi e^{-S[\chi] - \Delta S_k[\chi] + J^T \chi}, \quad (28)$$

$$W_k[J] = \log Z_k[J], \quad (29)$$

where  $J$  is the external field. As  $\Delta S_k[\chi]$  is large for low-wavenumber modes, the term  $\exp(-\Delta S_k[\chi])$  has the effect of focusing the integration on the high-wavenumber modes in  $\chi$ .

Applying the Legendre transformation to switch the independent variable  $J$  to  $\phi$ , we obtain the effective average action [7]:

$$\Gamma_k[\phi] \equiv \sup_J \{-W_k[J] + J^T \phi\} - \Delta S_k[\phi]. \quad (30)$$

Note that this transform should have the additional term  $-\Delta S_k[\phi]$  for the following reason [7]. From the supremum condition, we find:

$$\phi = \left( \frac{\delta W_k}{\delta J}[J] \right)^T = \langle \chi \rangle_{k,J} \equiv \frac{\int d\chi \chi e^{-S[\chi] - \Delta S_k[\chi] + J^T \chi}}{\int d\chi e^{-S[\chi] - \Delta S_k[\chi] + J^T \chi}} \quad (31)$$

$$= \frac{\int d\chi (\phi + \chi) e^{-S[\phi + \chi] - \Delta S_k[\chi] + (J^T - \phi^T R_k) \chi}}{\int d\chi e^{-S[\phi + \chi] - \Delta S_k[\chi] + (J^T - \phi^T R_k) \chi}}. \quad (32)$$

Eq.(31) appears to indicate that  $\phi$  is analogous to the conditional expectation of the control variable in the vicinity of  $\chi = 0$  under the existence of the external field  $J$ . However, if we insert into Eq.(32) the relation:

$$\frac{\delta \Gamma_k}{\delta \phi}[\phi] = J^T - \frac{\delta \Delta S_k}{\delta \phi}[\phi], \quad (33)$$

which is derived by the same operation as in Eq.(11), we obtain

$$\phi = \frac{\int d\chi (\phi + \chi) e^{-S[\phi + \chi] - \Delta S_k[\chi] + \frac{\delta \Gamma_k}{\delta \phi}[\phi] \chi}}{\int d\chi e^{-S[\phi + \chi] - \Delta S_k[\chi] + \frac{\delta \Gamma_k}{\delta \phi}[\phi] \chi}}. \quad (34)$$

This shows that  $\phi$  is in fact analogous to the conditional expectation of the control variable in the vicinity of  $\phi$  itself under the existence of the external field  $\delta \Gamma_k[\phi]/\delta \phi$ . Thus, the term  $-\Delta S_k[\phi]$  in Eq.(30) ensures that  $\phi$  is always the average of the surrounding  $\chi$ .

From Eq.(31) and the derivative of Eq.(33) with respect to  $\phi$ , we can see that the effective average action  $\Gamma_k[\phi]$  also satisfies the following equality:

$$\frac{\delta^2 \Gamma_k}{\delta \phi^2}[\phi] = \frac{\delta J}{\delta \phi} - \frac{\delta^2 \Delta S_k}{\delta \phi^2} \quad (35)$$

$$= \left( \frac{\delta^2 W_k}{\delta J^2}[J] \right)^{-1} - \frac{\delta^2 \Delta S_k}{\delta \phi^2} \quad (36)$$

$$= \left( \langle \chi \chi^T \rangle_{k,J} - \phi \phi^T \right)^{-1} - \frac{\delta^2 \Delta S_k}{\delta \phi^2}. \quad (37)$$

Eliminating  $J$  from Eq.(30) using Eq.(33), we have

$$\Gamma_k[\phi] = -W \left[ \frac{\delta \Gamma_k}{\delta \phi}[\phi] + \frac{\delta \Delta S_k}{\delta \phi}[\phi] \right] + \left( \frac{\delta \Gamma_k}{\delta \phi}[\phi] + \frac{\delta \Delta S_k}{\delta \phi}[\phi] \right) \phi - \Delta S_k[\phi] \quad (38)$$

$$= -\log \left\{ \int d\chi e^{-S[\chi] + \frac{\delta \Gamma_k}{\delta \phi}[\phi](\chi - \phi) - \Delta S_k[\chi - \phi]} \right\} \quad (39)$$

$$= -\log \left\{ \int d\chi e^{-S[\phi + \chi] + \frac{\delta \Gamma_k}{\delta \phi}[\phi]\chi - \Delta S_k[\chi]} \right\}. \quad (40)$$

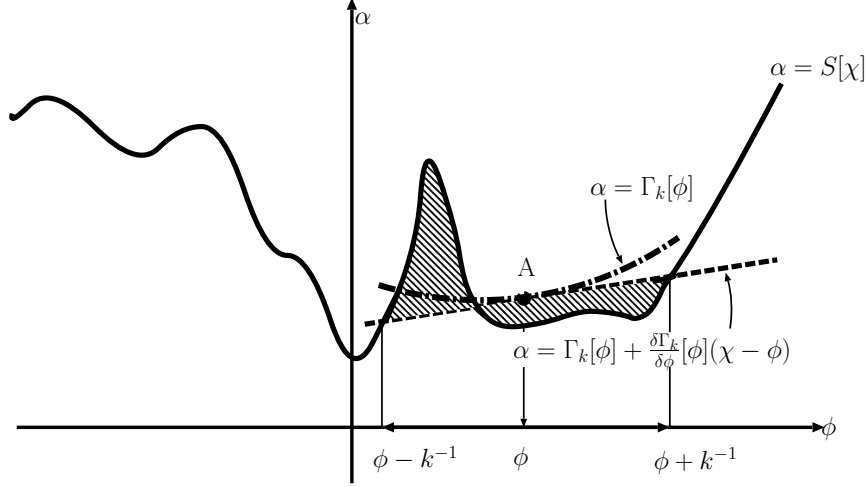


FIG. 1. The concept of the effective average action  $\Gamma_k$  (dotted-dashed curve). The tangent point  $A(\phi, \Gamma_k[\phi])$  and the slope  $\delta\Gamma_k[\phi]/\delta\phi$  are such that the point  $\phi$  coincides with the weighted average of the points  $\chi$  around the interval  $\phi - k^{-1} \leq \chi \leq \phi + k^{-1}$ . The weight exponentiates the deviation (shaded region) of the action  $\alpha = S[\chi]$  (solid curve) from the tangent plane  $\alpha = \Gamma_k[\phi] + (\delta\Gamma_k[\phi]/\delta\phi)(\chi - \phi)$  (dashed line). For simplicity, it is assumed that  $\chi$  is one-dimensional and the filter has  $R_k = k^2$ .

Although this has a recursive form about  $\Gamma_k[\phi]$ , we can write an approximation in closed form (see Appendix A 1 for the derivation):

$$\Gamma_k[\phi] \simeq S[\phi] + \frac{1}{2} \log \det \left\{ \frac{\delta^2}{\delta\phi^2} (S + \Delta S_k) [\phi] \right\}. \quad (41)$$

Note that we will not resort to such perturbation expansions in our numerical calculation, because it requires higher derivatives of the action, which are not always easy to calculate.

Fig. 1 illustrates the relation between the action and the effective average action in a simple case, where we assume  $\chi \in \mathbb{R}^1$  and the filter is  $R_k = k^2$ . The tangent point  $A(\phi, \Gamma_k[\phi])$  and the slope  $\delta\Gamma_k[\phi]/\delta\phi$  are such that the point  $\phi$  coincides with the weighted average of the points  $\chi$  around the interval  $\phi - k^{-1} \leq \chi \leq \phi + k^{-1}$ . Indeed, we see from Eqs. (34) and (40) that

$$\phi = \int d\chi \chi e^{-S[\chi] + \Gamma_k[\phi] + \frac{\delta\Gamma_k[\phi]}{\delta\phi}(\chi - \phi) - \frac{k^2}{2}(\chi - \phi)^2}. \quad (42)$$

One interesting thing about this smoothing is that it is not an averaging of the value of  $S[\chi]$  but of its independent variable  $\chi$ . Thereby, the effective average action serves as a kind of smoothed version of the cost function.

Taking Eqs. (26) and (27) into account,  $k \rightarrow \infty$  implies  $e^{-\Delta S_k[\chi]} = e^{-\frac{1}{2}\chi^T R_k \chi} \rightarrow \delta[\chi]$ , which leads Eq. (40) to

$$\Gamma_{k \rightarrow \infty}[\phi] = S[\phi]. \quad (43)$$

This means that when  $k$  is sufficiently large no fields around  $\phi$  are counted in  $\Gamma_k[\phi]$ , other than the field  $\phi$  itself, and thus the effective average action approaches the original action, or the cost function. In contrast, it is apparent that

$$\Gamma_{k \rightarrow 0}[\phi] = \Gamma[\phi]. \quad (44)$$

Hence, we have confirmed that  $\Gamma_k[\phi]$  constitutes a one-parameter family of functionals interpolating between the action  $S[\phi]$  and the effective action  $\Gamma[\phi]$ .

Hereafter, we assume that  $k$  has a finite value so as to integrate out some modes.

## B. Property of the stationary point

We assume that the stationary problem:

$$\frac{\delta\Gamma_k}{\delta\phi}[\phi] = J^T - \frac{\delta\Delta S_k}{\delta\phi}[\phi] = 0 \quad (45)$$

has stationary values at  $\hat{\phi}$ . From definition (31), we have:

$$\hat{\phi} = \langle \chi \rangle_{k, \frac{\delta \Delta S_k}{\delta \phi}[\hat{\phi}]} . \quad (46)$$

Since Eq. (45) can be thought of as the statistical equation of motion for the field  $\phi$ , the solution  $\hat{\phi}$  offers the estimated path for the statistical problem. At the stationary point, Eq. (32) reads:

$$\hat{\phi} = \frac{\int d\chi (\hat{\phi} + \chi) e^{-S[\hat{\phi} + \chi] - \Delta S_k[\chi]}}{\int d\chi e^{-S[\hat{\phi} + \chi] - \Delta S_k[\chi]}} . \quad (47)$$

This means that the stationary value  $\hat{\phi}$  provides the average with respect to high-wavenumber modes (see also Appendix A 2).

### III. ESTIMATION OF THE EFFECTIVE GRADIENT

#### A. Definition as an expected value

Rewriting Eq. (40), we introduce the exponent  $\Re[\phi, \chi]$  for convenience:

$$e^{-\Gamma_k[\phi]} = e^{-S[\phi]} \int d\chi e^{-\Re[\phi, \chi]}, \quad (48)$$

$$\Re[\phi, \chi] \equiv S[\phi + \chi] - S[\phi] - \frac{\delta \Gamma_k}{\delta \phi}[\phi] \chi + \Delta S_k[\chi], \quad (49)$$

$$\Gamma_k[\phi] = S[\phi] - \log \int d\chi e^{-\Re[\phi, \chi]}. \quad (50)$$

The gradient of  $\Gamma_k[\phi]$  is derived as the expected value  $\langle \cdot \rangle_{\Re}$  under the weight  $e^{-\Re}$ :

$$\begin{aligned} \frac{\delta \Gamma_k}{\delta \phi}[\phi] &= \frac{\delta S}{\delta \phi}[\phi] - \frac{\int d\chi \left( -\frac{\delta \Re}{\delta \phi}[\phi, \chi] \right) e^{-\Re[\phi, \chi]}}{\int d\chi e^{-\Re[\phi, \chi]}} \\ &= \left\langle \frac{\delta S}{\delta \phi}[\phi + \chi] \right\rangle_{\Re} - \langle \chi \rangle_{\Re}^T \frac{\delta^2 \Gamma_k}{\delta \phi^2}[\phi]. \end{aligned} \quad (51)$$

From Eq. (34), we have that

$$0 = \int d\chi \chi e^{-\Re[\phi, \chi]} \propto \langle \chi \rangle_{\Re}. \quad (52)$$

Thus, Eq. (51) can be simplified to

$$\frac{\delta \Gamma_k}{\delta \phi}[\phi] = \left\langle \frac{\delta S}{\delta \phi}[\phi + \chi] \right\rangle_{\Re}. \quad (53)$$

Eqs. (49) and (53) are recursive with respect to  $\delta \Gamma_k[\phi]/\delta \phi$ . Therefore, we need some approximation to enable a numerical evaluation. We may replace  $\delta \Gamma_k[\phi]/\delta \phi$  with the approximation  $\delta S[\phi]/\delta \phi$  in  $\Re[\phi, \chi]$  on the right-hand side of the equation. Eq. (53) can then be evaluated using the Metropolis method [9, 10]. We can then apply a successive correction procedure by updating the expectation with the latest value of  $\delta \Gamma_k[\phi]/\delta \phi$  in the weight.

Furthermore, from Eqs. (31), (34), and (37), we find that the second derivative can also be derived as the expected value:

$$\frac{\delta^2 \Gamma_k}{\delta \phi^2}[\phi] = \langle \chi \chi^T \rangle_{\Re}^{-1} - \frac{\delta^2 \Delta S_k}{\delta \phi^2}. \quad (54)$$

### B. Evaluation through the Metropolis method

In contrast to the case of the effective action in Eq. (16), the filtering term  $\Delta S_k[\chi]$  in Eq. (49) has the effect of confining the weight  $\exp(-\Re[\phi, \chi])$  to a small region in the control space, as illustrated in Fig. 1. Owing to this, we may assume that the expected value will be efficiently evaluated by a Markov-chain Monte Carlo method, e.g., the Metropolis adjusted Langevin algorithm [11]. Using the fact that the Langevin equation:

$$d\chi_t = \frac{1}{2} \nabla \log f(\chi_t) dt + dW_t, \quad (W_t : \text{the Wiener process}) \quad (55)$$

has the invariant distribution  $\pi(\chi) \equiv f(\chi) / \int d\chi f(\chi)$ , we construct a Markov chain by discretizing the equation and applying an acceptance/rejection procedure.

At time step  $n$ , according to the weight:

$$f(\chi^{(n)}) = e^{-\Re[\phi, \chi^{(n)}]}, \quad (56)$$

we define a proposal normal distribution:

$$q(\chi | \chi^{(n)}) = \mathcal{N}\left(\chi^{(n)} - \frac{\sigma^2}{2} \nabla_{\chi^{(n)}} \Re[\phi, \chi^{(n)}], \sigma^2 I\right), \quad (57)$$

$$\nabla_{\chi^{(n)}} \Re[\phi, \chi^{(n)}] = \left( \frac{\delta \Re}{\delta \chi^{(n)}}[\phi, \chi^{(n)}] \right)^T, \quad (58)$$

$$\frac{\delta \Re}{\delta \chi^{(n)}}[\phi, \chi^{(n)}] = \frac{\delta S}{\delta \phi}[\phi + \chi^{(n)}] - \frac{\delta \Gamma_k}{\delta \phi}[\phi] + \frac{\delta \Delta S_k}{\delta \phi}[\chi^{(n)}] \quad (59)$$

$$\simeq \frac{\delta S}{\delta \phi}[\phi + \chi^{(n)}] - \frac{\delta S}{\delta \phi}[\phi] + \frac{\delta \Delta S_k}{\delta \phi}[\chi^{(n)}] \quad (60)$$

to generate a random field  $\chi^*$  that obeys  $q$ :

$$\chi^* = \chi^{(n)} - \frac{\sigma^2}{2} \nabla_{\chi^{(n)}} \Re[\phi, \chi^{(n)}] + \sigma \xi, \quad \xi \sim \mathcal{N}(0, I). \quad (61)$$

We then update  $\chi^{(n+1)} = \chi^*$  with the acceptance probability:

$$\rho(\chi^{(n)}, \chi^*) = \min\left(1, \frac{f(\chi^*)}{f(\chi^{(n)})} \frac{q(\chi^{(n)} | \chi^*)}{q(\chi^* | \chi^{(n)})}\right), \quad (62)$$

or retain  $\chi^{(n+1)} = \chi^{(n)}$ . The ensemble of sample sequences  $\chi^{(n)}$  drawn in this way approximately follows the invariant distribution  $\pi(\chi)$ , and we can estimate the expected value accordingly:

$$\left\langle \frac{\delta S}{\delta \phi}[\phi + \chi] \right\rangle_{\Re} \simeq \frac{1}{N} \sum_{n=1}^N \frac{\delta S}{\delta \phi}[\phi + \chi^{(n)}]. \quad (63)$$

Note that, after the averaging, we may perform another refined averaging by substituting the derived sensitivity into the weight, because the term  $(\delta S[\phi]/\delta \phi)^T$  in Eq. (60) should have been  $(\delta \Gamma_k[\phi]/\delta \phi)^T$ .

Furthermore, within the limit of the accuracy of importance sampling, the finite difference of  $\Gamma_k$  can also be estimated as

$$\Gamma_k[\phi + \Delta \phi] - \Gamma_k[\phi] = -\log \left\langle e^{-\Re[\phi + \Delta \phi, \chi] + \Re[\phi, \chi]} \right\rangle_{\Re}. \quad (64)$$

### IV. A SIMPLE EXAMPLE

As a simple example, we consider a double-well potential:

$$S[\phi] = \frac{1}{2} (\phi^2 - a^2)^2, \quad a = 0.5, \quad R_k = 2^2. \quad (65)$$



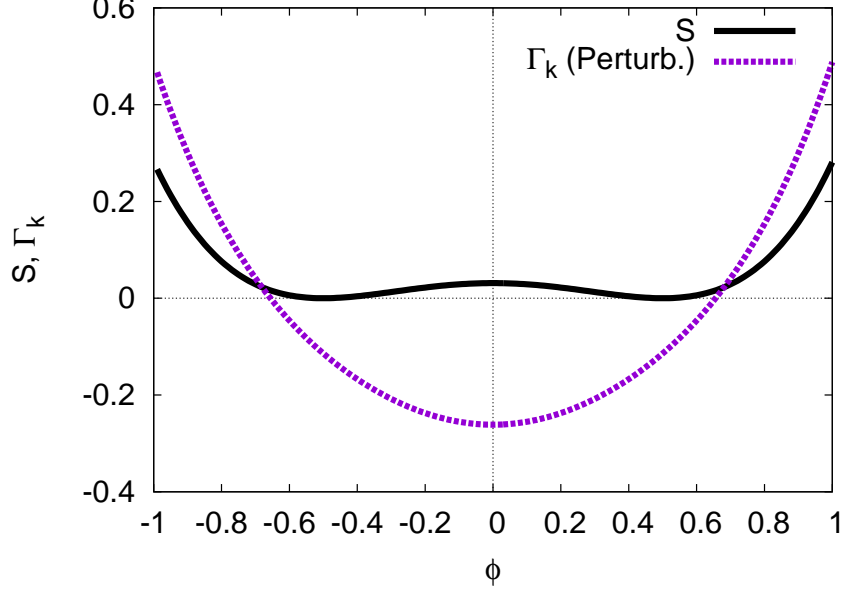


FIG. 2. Action  $S[\phi]$  (black) and effective average action  $\Gamma_k[\phi]$  evaluated using a perturbation expansion (purple) for the double-well potential

The effective average action given by a perturbation expansion up to the second order (see Appendix A 1) is:

$$\Gamma[\phi] = S[\phi] + \frac{1}{2} \log(6\phi^2 - 2a^2 + R_k). \quad (66)$$

Fig. 2 shows that the potential barrier at the center (black) is eliminated (purple) by integrating out the fluctuation. The gradients, along with the effective gradient of the action evaluated by the Metropolis method, are shown in Fig. 3. These curves suggest that the variational method using the coarse-grained sensitivity can capture the expected value  $\hat{\phi} = 0$ , which traditional variational methods will fail to find. This is because, in principle, variational methods are all designed to find one of the stationary values of the cost function, in this case  $\phi = \pm a$ .

## V. APPLICATION TO DATA ASSIMILATION

### A. Quadratic cost function

The cost function used in data assimilation is usually a quadratic in the nonlinear functional  $F[\phi]$ , such as [12]:

$$S[\phi] \equiv \frac{1}{2} F[\phi]^T F[\phi], \quad \frac{\delta S}{\delta \phi}[\phi] = F[\phi]^T \frac{\delta F}{\delta \phi}[\phi]. \quad (67)$$

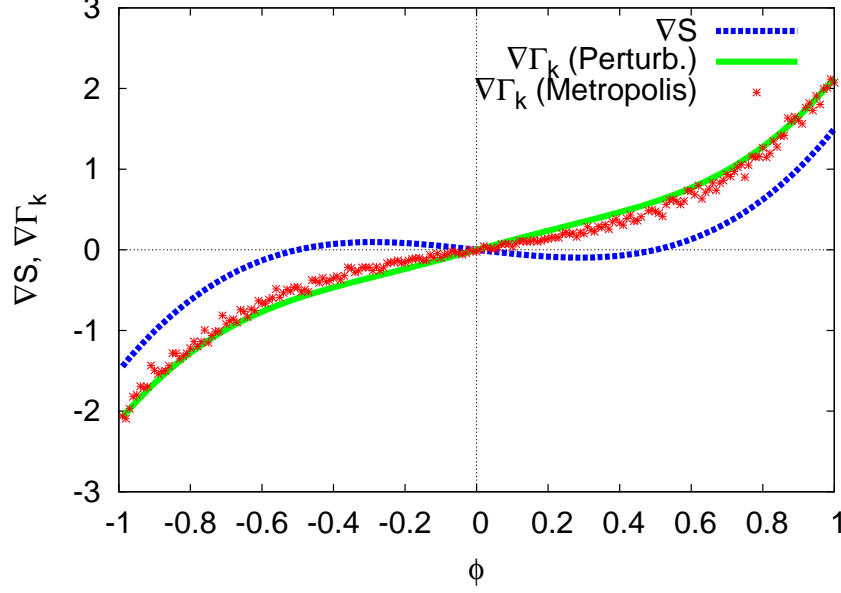


FIG. 3. Gradient of action  $\delta S[\phi]/\delta\phi$  (blue), effective gradient of action  $\delta\Gamma_k[\phi]/\delta\phi$  evaluated using a perturbation expansion (green), and effective gradient of action evaluated by the Metropolis method (red) for the double-well potential

To obtain the averaged sensitivity  $\delta\Gamma_k[\phi]/\delta\phi = \langle \delta S[\phi + \chi]/\delta\phi \rangle_{\Re}$ , we perform the following calculations:

$$\left( \frac{\delta S}{\delta\phi}[\phi + \chi] \right)^T = \left( \frac{\delta F}{\delta\phi}[\phi + \chi] \right)^T F[\phi + \chi], \quad (68)$$

$$\begin{aligned} \Re[\phi, \chi] &= \frac{1}{2} F[\phi + \chi]^T F[\phi + \chi] - \frac{1}{2} F[\phi]^T F[\phi] \\ &\quad - \chi^T \left( \frac{\delta\Gamma_k}{\delta\phi}[\phi] \right)^T + \Delta S_k[\chi] \end{aligned} \quad (69)$$

$$\begin{aligned} &\simeq \frac{1}{2} F[\phi + \chi]^T F[\phi + \chi] - \frac{1}{2} F[\phi]^T F[\phi] \\ &\quad - \chi^T \left( \frac{\delta F}{\delta\phi}[\phi] \right)^T F[\phi] + \Delta S_k[\chi], \end{aligned} \quad (70)$$

$$\begin{aligned} \nabla_{\chi} \Re[\phi, \chi] &= \left( \frac{\delta F}{\delta\phi}[\phi + \chi] \right)^T F[\phi + \chi] - \left( \frac{\delta\Gamma_k}{\delta\phi}[\phi] \right)^T \\ &\quad + \left( \frac{\delta\Delta S_k}{\delta\chi}[\chi] \right)^T. \end{aligned} \quad (71)$$

Thus, each sample requires a forward integration  $F[\phi + \chi]$  and a subsequent adjoint integration  $(\delta F[\phi + \chi]/\delta\phi)^T$ .

### B. The Logistic map

We now consider a smoothing problem for the Logistic map  $\Psi \mapsto r\Psi(1 - \Psi)$  [13, 14] fitted to observation  $y$ . As a data assimilation problem, we use the following cost function  $S$  and its gradient[15]:

$$S[\phi] = \frac{1}{2\sigma_0^2}(\phi - m_0)^2 + \sum_{j=0}^{J-1} \frac{1}{2\gamma^2} \left( y_{j+1} - \Psi^{(j+1)}[\phi] \right)^2, \quad (72)$$

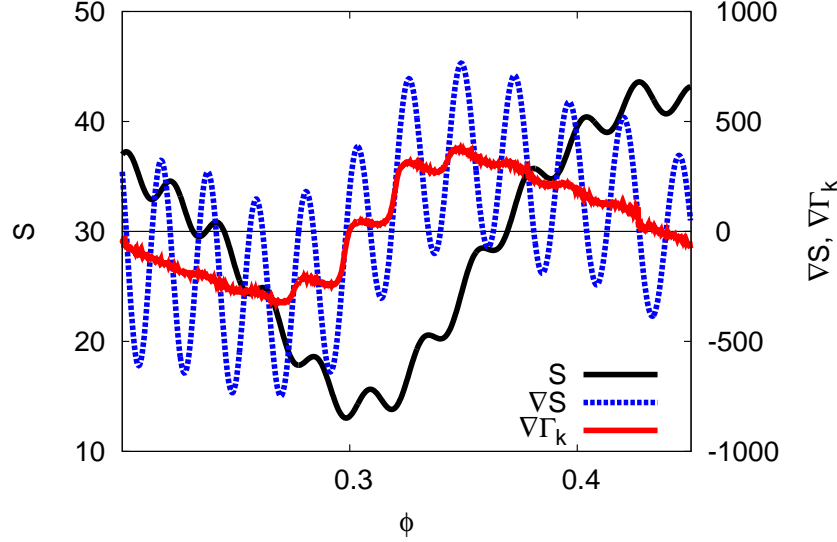


FIG. 4. Action  $S[\phi]$  (black), gradient of action  $\delta S[\phi]/\delta\phi$  (blue), and effective gradient of action  $\delta\Gamma_k[\phi]/\delta\phi$  (red) for the Logistic map. The true value for the data assimilation problem is  $\phi = 0.3$ , and the first-guess is 0.4.

$$\Psi^{(j+1)}[\phi] = r\Psi^{(j)}[\phi] \left(1 - \Psi^{(j)}[\phi]\right), \quad \Psi^{(0)}[\phi] = \phi, \quad (73)$$

$$\begin{aligned} \frac{\delta S}{\delta\phi}[\phi] &= \frac{1}{\sigma_0^2}(\phi - m_0) + \frac{1}{\gamma^2} \sum_{j=0}^{J-1} \left\{ \prod_{l=0}^j r(1 - 2\Psi^{(l)}[\phi]) \right\} \\ &\times \left( \Psi^{(j+1)}[\phi] - y_{j+1} \right), \end{aligned} \quad (74)$$

where  $\sigma_0^2$  and  $\gamma^2$  are the background and observational error variances, respectively. The parameters are set to  $r = 4$ ,  $J = 6$ ,  $\sigma_0 = 0.1$ ,  $\gamma = 0.2$ , and  $m_0 = 0.4$  (first-guess), similar to those in [14]. The observations are sampled from a model sequence given by the initial value  $v_0 = 0.3$  added to observational noise.

The optimization problem can apparently be solved using a variational method that seeks the optimal initial condition using the gradient information. However, there are multiple extrema of the cost function (see black curve in Fig. 4), which makes it difficult to find the global minimum. Thus, we should utilize the effective gradient (63), derived using the Metropolis method. We use a finite constant  $R_k = (0.008)^{-2}$ , where 0.008 is the typical half-wavelength of short fluctuations in  $S[\phi]$ .

Fig. 4 shows the action (black), the gradient of action (blue), and the effective gradient of action (red) for this system. It is clear that the original gradient has too many zeros for worthwhile variational data assimilation using the gradient. However, because the effective gradient has relatively few zeros, it can be applied to a variational data assimilation to find the minimum at around  $\phi = v_0$ , as long as the first-guess is not far from the true value.

We performed two data assimilation experiments using the steepest descent method with the gradient  $\delta S[\phi]/\delta\phi$  and the effective gradient  $\delta\Gamma_k[\phi]/\delta\phi$  (see the algorithm in Appendix B). As shown in Fig. 5, the case with  $\delta S[\phi]/\delta\phi$  converges to a local minimum  $\phi = 0.388$ , whereas the case with  $\delta\Gamma_k[\phi]/\delta\phi$  converges to  $\phi = 0.315$  around the global minimum, which indicates the superiority of the effective gradient. Note that the decrease in the cost function  $S[\phi]$  in the latter case is not necessarily monotonic, because the optimization problem is actually defined for the effective average action  $\Gamma_k[\phi]$ .

This illustrates the potential usefulness of the effective gradient (sensitivity) for data assimilation.

### C. Lorenz '96 model

Next, we examine the sensitivity that appears during data assimilation in the Lorenz '96 model, which is a simple dynamical system designed to mimic the dynamics of Rossby waves in atmospheric dynamics [14, 16]. This system

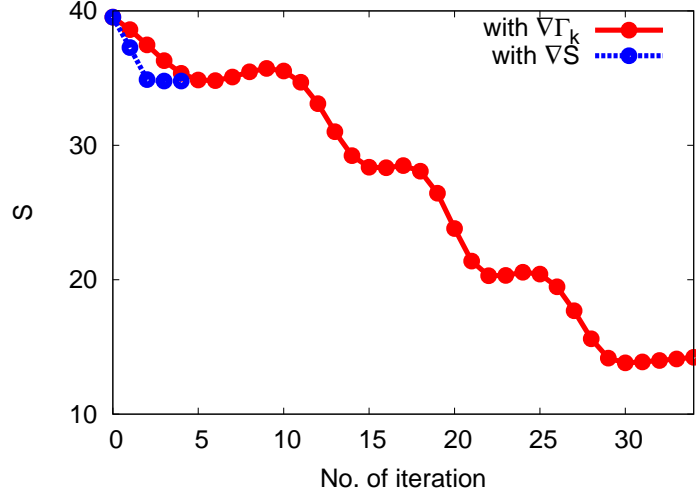


FIG. 5. The variation of the cost function during data assimilation using the steepest descent method with  $\delta S[\phi]/\delta\phi$  (blue) and  $\delta\Gamma_k[\phi]/\delta\phi$  (red). The former converges to a local minimum  $\phi = 0.388$ ; the latter converges to  $\phi = 0.315$ , which is close to the global minimum.

can be written as:

$$\frac{d\theta_l}{dt} = \theta_{l-1}(\theta_{l+1} - \theta_{l-2}) - \theta_l + F, \quad l = 1, 2, \dots, M, \quad (75)$$

$$\theta_0 = \theta_M, \quad \theta_{M+1} = \theta_1, \quad \theta_{-1} = \theta_{M-1}. \quad (76)$$

We write the time evolution operator from the initial condition as:

$$\Psi^{(j)}[\theta(t=0)] = \theta(t=t_j), \quad (77)$$

with time-step  $t_{j+1} - t_j = \Delta t$ . If we choose the initial condition  $\theta(t=0) = \phi \in \mathbb{R}^M$  as the control variable, we can define a similar cost function to that in Eq. (72):

$$S[\phi] = \frac{1}{2\sigma_0^2} \|\phi - m_0\|^2 + \sum_{j=0}^{J-1} \frac{1}{2\gamma^2} \|y_{j+1} - \Psi^{(j+1)}[\phi]\|^2, \quad (78)$$

where  $\|\cdot\|$  is the Euclidean norm. The parameters are set to  $F = 6$ ,  $J = 32$ ,  $M = 20$ ,  $\sigma_0 = 2$ , and  $\gamma = 2$ . With these parameters, this model is in an unstable regime with the first Lyapunov exponent  $\lambda_1 \simeq 0.84 > 0$ . The true initial condition  $v_0 \in \mathbb{R}^M$  is given by model integration within an interval from a randomly chosen initial condition. The observation is sampled from a model sequence starting from  $v_0$ , with observational noise added to the sample. The observation is defined only at the times  $j+1 = 3, 6, \dots, 30$  and space  $l = 1, 2, \dots, 8$ . The first-guess  $m_0$  is given by changing only the first component of  $v_0$ :

$$(m_0)_1 = (v_0)_1 + \sigma_0, \quad (m_0)_2 = (v_0)_2, \dots, (m_0)_M = (v_0)_M. \quad (79)$$

The filter  $R_k$  is in the form of Eq. (24) with  $j_k = 10$  and  $k = 0.25^{-1}$ .  $k$  is set so that the filtering term  $\Delta S_k[\phi]$  is of order 1, that is,  $O(1) = \Delta S_k[\phi] \simeq k^2 \text{var}(\phi)$ , where the typical fluctuation is assumed to be  $\text{var}(\phi) \simeq k^{-2} = 0.25^2$ . The time evolution of (75) and its adjoint are solved by the Runge-Kutta method with time-step  $\Delta t = 0.1$ . The experiments are designed to investigate how the action, the gradient of action, and the effective gradient of action change if we move the control variable as:

$$\phi_1 = (v_0)_1 + \sigma_0 \eta, \quad \phi_2 = (v_0)_2, \dots, \phi_M = (v_0)_M, \quad (80)$$

where  $-1 \leq \eta \leq 1$ .

Fig. 6 shows the action (black), the gradient of action (blue), and the effective gradient of action (red) for this experiment. It is clear that the original gradient has several zeros that will complicate the variational data assimilation

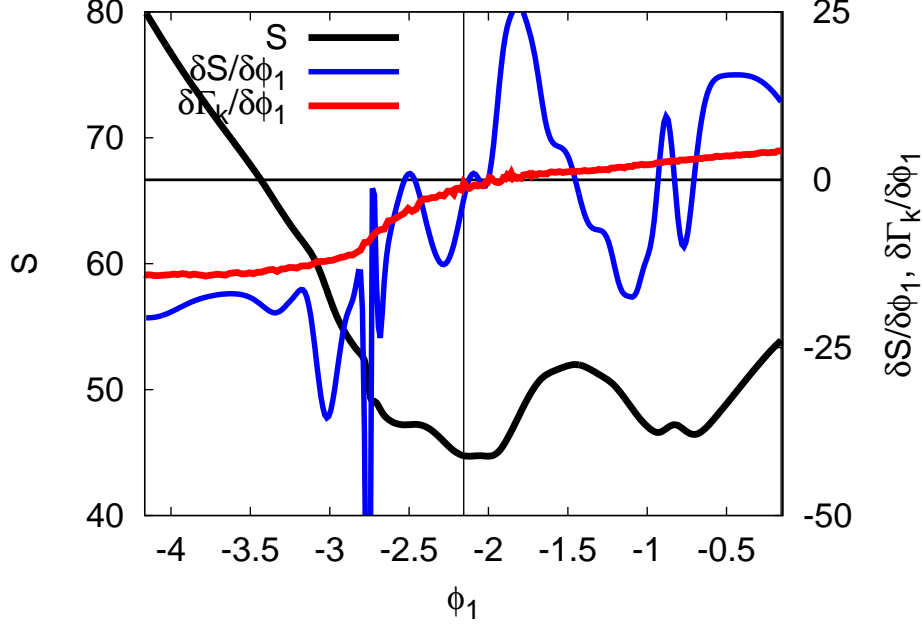


FIG. 6. Action  $S[\phi]$  (black), gradient of action  $\delta S[\phi]/\delta\phi_1$  (blue), and effective gradient of action  $\delta\Gamma_k[\phi]/\delta\phi_1$  (red) for the Lorenz '96 model. The true value for the data assimilation problem is  $\phi_1 = -2.156$ , and the first-guess is  $-0.156$ .

using the gradient. However, the effective gradient has only one zero near the true value, which is a similar result as for the Logistic map.

Thus, this example also shows the potential usefulness of the effective gradient for finding the optimal initial condition.

#### D. Slow and fast degrees of freedom

Toward the application to high-dimensional systems, we briefly note a possible procedure for treating two distinct spatial modes in the coarse-grained data assimilation.

Assume the infra-red filter has a sharp cutoff in the momentum representation:

$$R_k(p) = \begin{cases} a^2 & \text{if } |p| < k \\ 0 & \text{otherwise} \end{cases}, \quad (81)$$

and the fluctuation  $\chi$  can be decomposed into:

$$\chi(p) = \begin{cases} \chi_s(p) & \text{if } |p| < k \\ \chi_f(p) & \text{otherwise} \end{cases}. \quad (82)$$

Then, as the “mass”  $a^2 \rightarrow \infty$ , we have:

$$e^{-\Delta S_k[\chi_s + \chi_f]} = e^{-\frac{a^2}{2} \chi_s^T \chi_s} \rightarrow \delta[\chi_s], \quad (83)$$

which leads to:

$$\frac{\delta\Gamma_k}{\delta\phi}[\phi] = \frac{\int d\chi_s \int d\chi_f \frac{\delta S}{\delta\phi}[\phi + \chi_s + \chi_f] e^{-S[\phi + \chi_s + \chi_f] + S[\phi] + \frac{\delta\Gamma_k}{\delta\phi}[\phi](\chi_s + \chi_f)} \delta[\chi_s]}{\int d\chi_s \int d\chi_f e^{-S[\phi + \chi_s + \chi_f] + S[\phi] + \frac{\delta\Gamma_k}{\delta\phi}[\phi](\chi_s + \chi_f)} \delta[\chi_s]} \quad (84)$$

$$\rightarrow \frac{\int d\chi_f \frac{\delta S}{\delta\phi}[\phi + \chi_f] e^{-S[\phi + \chi_f] + S[\phi] + \frac{\delta\Gamma_k}{\delta\phi}[\phi]\chi_f}}{\int d\chi_f e^{-S[\phi + \chi_f] + S[\phi] + \frac{\delta\Gamma_k}{\delta\phi}[\phi]\chi_f}}. \quad (85)$$

Hence, the sensitivity can be calculated as the average under the weight that integrates out the fast degrees of freedom, which will contribute to reducing the dimensionality of the path space. In the case of coupled atmosphere–ocean systems, we can assume that the atmospheric system is represented by  $\chi_f$  and the oceanic system is  $\chi_s$ . This suggests that, through this coarse-graining procedure, the sensitivity regarding the coupled system can be expressed by the slow oceanic variables alone.

### E. Two-scale Lorenz '96 model

To illustrate the application to multiscale systems, as mentioned in Section V D, we examine the sensitivity appearing during data assimilation in the two-scale Lorenz '96 model [16]. This system can be written for the slow ( $\zeta$ ) and the fast ( $\xi$ ) variables as:

$$\frac{d\zeta_{l_1}}{dt} = \zeta_{l_1-1} (\zeta_{l_1+1} - \zeta_{l_1-2}) - \zeta_{l_1} + F - \frac{hc}{b} \sum_{l_2=1}^{M_2} \xi_{l_2, l_1}, \quad l_1 = 1, 2, \dots, M_1, \quad (86)$$

$$\frac{d\xi_{l_2, l_1}}{dt} = cb\xi_{l_2+1, l_1} (\xi_{l_2-1, l_1} - \xi_{l_2+2, l_1}) - c\xi_{l_2, l_1} + \frac{hc}{b}\zeta_{l_1}, \quad l_2 = 1, 2, \dots, M_2, \quad l_1 = 1, 2, \dots, M_1, \quad (87)$$

$$\zeta_0 = \zeta_{M_1}, \quad \zeta_{M_1+1} = \zeta_1, \quad \zeta_{-1} = \zeta_{M_1-1}, \quad (88)$$

$$\xi_{0, l_1} = \xi_{M_2, l_1-1}, \quad \xi_{M_2+1, l_1} = \xi_{1, l_1+1}, \quad \xi_{M_2+2, l_1} = \xi_{2, l_1+1}, \quad l_1 = 1, 2, \dots, M_1, \quad (89)$$

where  $\zeta(t) \in \mathbb{R}^{M_1}$ ,  $\xi(t) \in \mathbb{R}^{M_1 M_2}$ . We write the time evolution operator of the state  $\theta = (\zeta^T, \xi^T)^T$  from the initial condition as:

$$\Psi^{(j)}[\theta(t=0)] = \theta(t=t_j), \quad (90)$$

with time-step  $t_{j+1} - t_j = \Delta t$ . If we choose the initial condition  $\theta(t=0) = \phi \in \mathbb{R}^M$ , where  $M = M_1 + M_1 M_2$ , as the control variable, we can define a cost function as follows:

$$S[\phi] = \frac{1}{2\sigma_0^2} \|\phi - m_0\|^2 + \sum_{j=0}^{J-1} \frac{1}{2\gamma^2} \left\| y_{j+1} - \Psi^{(j+1)}[\phi] \right\|^2, \quad (91)$$

where  $\|\cdot\|$  is the Euclidean norm. The parameters are set to  $F = 6$ ,  $J = 110$ ,  $M_1 = 5$ ,  $M_2 = 3$ ,  $h = 1.6$ ,  $b = c = 10$ ,  $\sigma_0 = 2$ , and  $\gamma = 2$ . With these parameters, this model is in an unstable regime with the first Lyapunov exponent  $\lambda_1 \simeq 9.0 > 0$ . The true initial condition  $v_0 \in \mathbb{R}^M$  is given by model integration within an interval from a randomly chosen initial condition. The observation is sampled from a model sequence starting from  $v_0$ , with observational noise added to the sample. The observation is defined only at the times  $j+1 = 3, 6, \dots, 108$  and for the slow variables  $l_1 = 1, 2, \dots, M_1$ . The first-guess  $m_0$  is given by changing only the first component of  $v_0$ :

$$(m_0)_1 = (v_0)_1 + \sigma_0, \quad (m_0)_2 = (v_0)_2, \dots, (m_0)_M = (v_0)_M. \quad (92)$$

Similar to Eq. (81), we set the filter  $R_k$  as

$$R_k = \begin{cases} k_\zeta^2 & \text{for } \zeta(t=0) \\ k_\xi^2 & \text{for } \xi(t=0) \end{cases}, \quad (93)$$

with  $k_\zeta = 0.01^{-1}$  and  $k_\xi = 0.04^{-1}$ . This setting,  $k_\zeta > k_\xi$ , mainly integrates out the fast degrees of freedom  $\xi(t=0)$ , whose typical fluctuation is assumed to be  $\text{var}(\xi(t=0)) \simeq k_\xi^{-2} = 0.04^2$ . The time evolution of (86) and (87) as well as their adjoints are solved by the Runge–Kutta method with time-step  $\Delta t = 0.006$ . As in section V C, the experiments

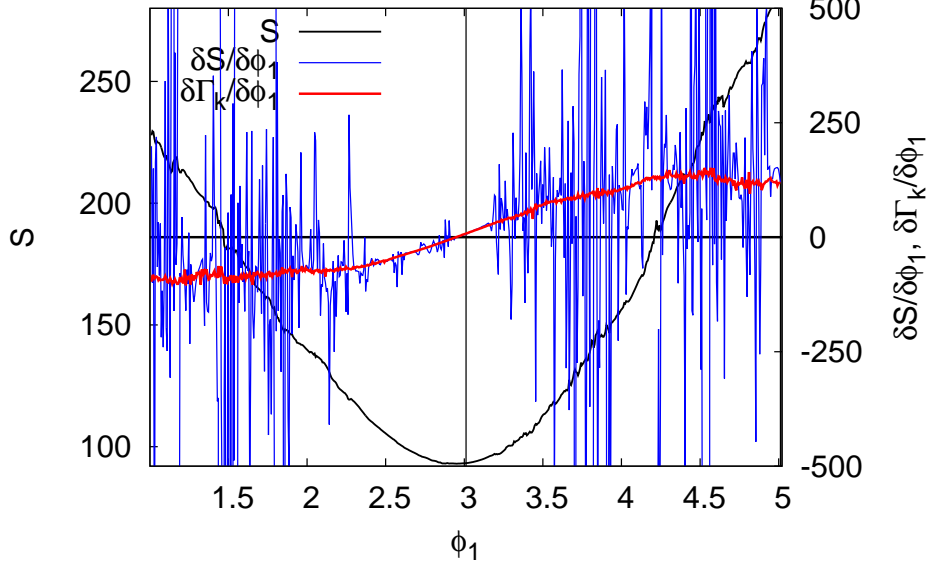


FIG. 7. Action  $S[\phi]$  (black), gradient of action  $\delta S[\phi]/\delta\phi_1$  (blue), and effective gradient of action  $\delta\Gamma_k[\phi]/\delta\phi_1$  (red) for the two-scale Lorenz '96 model. The true value for the data assimilation problem is  $\phi_1 = 3.011$ , and the first-guess is 5.011.

are designed to investigate the changes in the action, the gradient of action, and the effective gradient of action if we move the control variable as:

$$\phi_1 = (v_0)_1 + \sigma_0\eta, \phi_2 = (v_0)_2, \dots, \phi_M = (v_0)_M, \quad (94)$$

where  $-1 \leq \eta \leq 1$ .

Fig. 7 shows the action (black), the gradient of action (blue), and the effective gradient of action (red) for this experiment. The action is almost parabolic in shape with many small bumps due to the fast degrees of freedom. Consequently, the gradient has many zeros that can complicate the variational data assimilation using it. However, the effective gradient has only one zero near the true value, similar to the result obtained for the Logistic map and for the one-scale Lorenz '96 model. Thus, this example shows the potential application of the effective gradient for finding the optimal initial condition in multiscale data assimilation.

## VI. CONCLUSION

We have investigated the use of the effective average action and its gradient in multiscale data assimilation. A framework has been proposed that allows the multiscale data assimilation problem to be solved by replacing the cost function with the effective average action. We have also proposed a novel method of evaluating the gradient of the effective average action numerically. In principle, this enables a broader range of data assimilation problems to be solved by seeking the stationary point of this effective average action numerically using its gradient.

This work can be summarized as follows.

- The concept of the effective average action provides a consistent framework for data assimilation in nonlinear multiscale systems.
- If we can numerically evaluate the path integral:

$$\frac{\delta\Gamma_k}{\delta\phi}[\phi] = \frac{\int d\chi \frac{\delta S}{\delta\phi}[\phi + \chi] e^{-\Re[\phi, \chi]}}{\int d\chi e^{-\Re[\phi, \chi]}} \equiv \left\langle \frac{\delta S}{\delta\phi}[\phi + \chi] \right\rangle_{\Re},$$

$$\Re[\phi, \chi] \equiv S[\phi + \chi] - S[\phi] - \frac{\delta\Gamma_k}{\delta\phi}[\phi]\chi + \Delta S_k[\chi]$$

with reasonable accuracy and computational burden, then we obtain the coarse-grained sensitivity, which constitutes a key factor in the variational data assimilation of a coarse-grained field.

- The proposed procedure estimates the coarse-grained sensitivity through the Metropolis method by averaging an ensemble of original sensitivities that are distributed according to weights related to the nonlinearity:

$$\left\langle \frac{\delta S}{\delta \phi}[\phi + \chi] \right\rangle_{\mathfrak{R}} \simeq \frac{1}{N} \sum_{n=1}^N \frac{\delta S}{\delta \phi}[\phi + \chi^{(n)}].$$

- The stationary problem for the effective average action  $\Gamma_k[\phi]$  can be solved using a gradient method with the gradient  $\delta\Gamma_k[\phi]/\delta\phi$ .
- The stationary value  $\hat{\phi}$  represents the extremum for the coarse-grained field after integrating out the ultra-violet fluctuations. This can be regarded as a solution of the multiscale data assimilation problem.
- We demonstrated the usefulness of the effective gradient for data assimilation in a simple setting with the double-well potential, the Logistic map, the one-scale Lorenz '96 model, and the two-scale Lorenz '96 model.

Future research should consider the following issues.

- The infra-red filter works well when we can separate slow and fast modes cleanly, as described in sections V D and V E. However, in general, we have to deal with control variables that have continuous spectra. For such cases, we should carefully consider an infra-red filter design that is suitable for revealing the slow dynamics of the system under consideration.
- In the case of a larger system, the computational burden of the Metropolis method could be huge, because we require many samples to yield a statistically reasonable integration result. Moreover, this should be incorporated into a recursive procedure, or a fixed point calculation.

Despite these technical difficulties, the coarse-grained sensitivities are of great importance, since they provide an invaluable perspective on the slow dynamics of multiscale systems. It should be noted that our approach in the present form has a fairly limited scope of application to data assimilation problems in geoscience, which typically require more high-dimensional systems.

## ACKNOWLEDGMENTS

The author is grateful to H. D. I. Abarbanel for motivating this research and for several suggestions. The numerical simulations were performed on JAMSTEC SC supercomputer system.

## Appendix A: Properties of $\Gamma_k$

### 1. Perturbation expansion

To evaluate the path integral (50) through a perturbation expansion, we apply the approximation  $\delta\Gamma_k[\phi]/\delta\phi \simeq \delta S[\phi]/\delta\phi$  in the exponent of Eq. (49), and truncate the Taylor series expansion of  $S[\phi + \chi]$  to the quadratic order:

$$\begin{aligned} S[\phi + \chi] &= S[\phi] - \frac{\delta\Gamma_k}{\delta\phi}[\phi]\chi + \Delta S_k[\chi] \\ &\simeq S[\phi] + \frac{\delta S}{\delta\phi}[\phi]\chi + \frac{1}{2}\chi^T \frac{\delta^2 S}{\delta\phi^2}[\phi]\chi - S[\phi] - \frac{\delta S}{\delta\phi}[\phi]\chi + \frac{1}{2}\chi^T R_k \chi \\ &= \frac{1}{2}\chi^T \frac{\delta^2 S}{\delta\phi^2}[\phi]\chi + \frac{1}{2}\chi^T R_k \chi = \frac{1}{2}\chi^T \left( \frac{\delta^2 S}{\delta\phi^2}[\phi] + R_k \right) \chi. \end{aligned} \tag{A1}$$



Taking the Gaussian integral, we obtain

$$\begin{aligned}\Gamma_k[\phi] &\simeq S[\phi] - \log \int d\chi e^{-\frac{1}{2}\chi^T \left( \frac{\delta^2 S}{\delta \phi^2}[\phi] + R_k \right) \chi} \\ &= S[\phi] + \frac{1}{2} \log \det \left( \frac{\delta^2 S}{\delta \phi^2}[\phi] + R_k \right).\end{aligned}\tag{A2}$$

Thus, we require at least the second derivative of the action  $S[\phi]$  for the perturbation calculation of  $\Gamma_k[\phi]$ .

The same procedure can be applied to the gradient (53) as follows:

$$\begin{aligned}\frac{\delta \Gamma_k}{\delta \phi}[\phi] &\simeq \frac{\delta S}{\delta \phi}[\phi] + \frac{1}{2} \left\langle \chi^T \left( \frac{\delta^3 S}{\delta \phi^3}[\phi] \right) \chi \right\rangle_{\mathfrak{R}} \\ &= \frac{\delta S}{\delta \phi}[\phi] + \frac{1}{2} \text{tr} \left\{ \frac{\delta^3 S}{\delta \phi^3}[\phi] \left( \frac{\delta^2 S}{\delta \phi^2}[\phi] + R_k \right)^{-1} \right\}.\end{aligned}\tag{A3}$$

Hence, we require at least the second and third derivatives of the action  $S[\phi]$  for the perturbation calculation of the gradient of  $\Gamma_k[\phi]$ .

## 2. Relationship to the growth of instabilities

We consider the case where  $S[\phi]$  is the cost function of strong-constraint 4D-Var [17]. The meaning of  $\log \det$  in Eq. (A2) can be clarified by considering the basis of singular vectors. With the singular values  $\sigma_1 > \sigma_2 > \dots$ , we can write:

$$\frac{\delta^2 S}{\delta \phi^2}[\phi] \simeq \left( \frac{\delta F}{\delta \phi}[\phi] \right)^T \left( \frac{\delta F}{\delta \phi}[\phi] \right) = \text{diag}[\sigma_1^2, \sigma_2^2, \dots].\tag{A4}$$

Using a positive constant  $k \gg 1$ , we can define the infra-red filter as:

$$R_k = \begin{cases} k^2 & \text{if } \sigma_i < k \\ 0 & \text{otherwise} \end{cases}.\tag{A5}$$

Then, we have

$$\frac{\delta^2 S}{\delta \phi^2}[\phi] + R_k \simeq \text{diag}[\sigma_1^2, \sigma_2^2, \dots, k^2, \dots, k^2],\tag{A6}$$

$$\frac{1}{2} \log \det \left( \frac{\delta^2 S}{\delta \phi^2}[\phi] + R_k \right) \simeq \sum_{\sigma_i \geq k} \log \sigma_i + \sum_{\sigma_i < k} \log k.\tag{A7}$$

That is, the term  $\log \det$  represents the sum of the logarithms of the leading singular values. The additional term in Eq. (A7) has the effect of integrating out the growing disturbances in the cost function.

## Appendix B: An algorithm for coarse-grained data assimilation

---

**Algorithm 1** coarse-grained\_data\_assimilation
 

---

```

 $\Psi_0 \leftarrow v_0$  ▷ set true value
for  $t = 0 \rightarrow T - 1$  do
   $\Psi_{t+1} \leftarrow \psi(\Psi_t)$ 
  generate  $\xi \sim \mathcal{N}(0, I)$ 
   $y_{t+1} \leftarrow \Psi_{t+1} + \gamma \xi$  ▷ set observation
end for
 $\phi_0 \leftarrow m_0$  ▷ set first-guess
for  $i = 0 \rightarrow I - 1$  do ▷ assimilation loop
  CALC_COST( $y_{1:T}, \phi_i, \Psi_{0:T}, S_i$ )
  CALC_SENSITIVITY( $y_{1:T}, \Psi_{0:T}, \nabla S_i$ ) ▷ See gradient_calculations
  CALC_COARSE-GRAINED_SENSITIVITY( $y_{1:T}, \phi_i, S_i, \nabla S_i, \nabla \Gamma_i$ ) ▷ See gradient_calculations
  if  $|\nabla \Gamma_i| < C_{th}$  then
    return  $\phi_i, S_i$ 
  end if
   $\phi_{i+1} \leftarrow \phi_i - \alpha \nabla \Gamma_i$  ▷ update control variable
end for

procedure CALC_COST( $y_{1:T}, \phi_i, \Psi_{0:T}, S_i$ )
   $\Psi_0 \leftarrow \phi_i$  ▷ background term
   $S_i \leftarrow \frac{1}{2\sigma_0^2} |\Psi_0 - m_0|^2$ 
  for  $t = 0 \rightarrow T - 1$  do
     $\Psi_{t+1} \leftarrow \psi(\Psi_t)$  ▷ forward time stepping
     $S_i \leftarrow S_i + \frac{1}{2\gamma^2} |y_{t+1} - \Psi_{t+1}|^2$  ▷ observational term
  end for
end procedure

```

---

**Algorithm 2** gradient\_calculations

---

```

procedure CALC_SENSITIVITY( $y_{1:T}, \Psi_{0:T}, \nabla S_i$ )
   $\hat{\Psi}_{0:T} \leftarrow 0$ 
  for  $t = T - 1 \rightarrow 0$  do
     $\hat{\Psi}_{t+1} \leftarrow \hat{\Psi}_{t+1} - \frac{1}{\gamma^2}(y_{t+1} - \Psi_{t+1})$  ▷ observational term
     $\hat{\Psi}_t \leftarrow \hat{\Psi}_t + \left(\frac{\partial \psi}{\partial \Psi_t}\right)^T \hat{\Psi}_{t+1}$  ▷ adjoint time stepping
     $\hat{\Psi}_{t+1} \leftarrow 0$ 
  end for
   $\hat{\Psi}_0 \leftarrow \hat{\Psi}_0 + \frac{1}{\sigma_0^2}(\Psi_0 - m_0)$  ▷ background term
   $\nabla S_i \leftarrow \hat{\Psi}_0$ 
   $\hat{\Psi}_0 \leftarrow 0$ 
end procedure

procedure CALC_COARSE-GRAINED_SENSITIVITY( $y_{1:T}, \phi_i, S_i, \nabla S_i, \nabla \Gamma_i$ )
   $\nabla \Gamma_i \leftarrow \nabla S_i$  ▷ first-guess of gradient in weight
  for  $l = 1 \rightarrow L$  do ▷ successive correction of gradient in weight
     $\chi \leftarrow 0$ 
     $\nabla S \leftarrow \nabla S_i$ 
     $R \leftarrow 0$ 
     $\nabla R \leftarrow 0$ 
     $\nabla \Gamma^{\text{acc}} \leftarrow 0$ 
    for  $n = 0 \rightarrow N - 1$  do ▷ Markov-chain loop
      generate  $\xi \sim \mathcal{N}(0, I)$ 
       $\chi^* \leftarrow \chi - \frac{1}{2}\sigma^2 \nabla R + \sigma \xi$  ▷ proposed
      CALC_COST( $y_{1:T}, \phi_i + \chi^*, \Psi_{0:T}, S^*$ )
      CALC_SENSITIVITY( $y_{1:T}, \Psi_{0:T}, \nabla S^*$ )
       $R^* \leftarrow S^* - S_i - \langle \nabla \Gamma_i, \chi^* \rangle + \frac{1}{2}k^2 |\chi^*|^2$ 
       $\nabla R^* \leftarrow \nabla S^* - \nabla \Gamma_i + k^2 \chi^*$ 
       $q_+ \leftarrow \frac{1}{2\sigma^2} |\chi - \frac{1}{2}\sigma^2 \nabla R - \chi^*|^2$ 
       $q_- \leftarrow \frac{1}{2\sigma^2} |\chi^* - \frac{1}{2}\sigma^2 \nabla R^* - \chi|^2$ 
       $a \leftarrow -R^* + R - q_- + q_+$ 
      generate  $\zeta \sim \mathcal{U}(0, 1)$ 
      if  $a \geq 0$  or  $\zeta < \exp(a)$  then ▷ Metropolis criterion
         $\chi \leftarrow \chi^*$ 
         $\nabla S \leftarrow \nabla S^*$ 
         $R \leftarrow R^*$ 
         $\nabla R \leftarrow \nabla R^*$ 
      end if
       $\nabla \Gamma^{\text{acc}} \leftarrow \nabla \Gamma^{\text{acc}} + \nabla S$  ▷ accumulate gradient
    end for
     $\nabla \Gamma_i \leftarrow \nabla \Gamma^{\text{acc}} / N$  ▷ take average
  end for
end procedure

```

---

## REFERENCES

- 
- [1] M. Bocquet, L. Wu, and F. Chevallier, Quarterly Journal of the Royal Meteorological Society **137**, 1340 (2011).  
[2] M. Peña and E. Kalnay, Nonlinear Processes in Geophysics **11**, 319 (2004).  
[3] H. D. Abarbanel, Physics Letters A **373**, 4044 (2009).  
[4] H. D. Abarbanel, *Predicting the Future: Completing Models of Observed Complex Systems* (Springer, 2013).  
[5] The convexity of  $W[J]$  follows from Hölder's inequality:  $E[\exp((1-\gamma)J_1^T \chi + \gamma J_2^T \chi)] \leq E[\exp(J_1^T \chi)]^{1-\gamma} E[\exp(J_2^T \chi)]^\gamma$ ,  $0 \leq \gamma \leq 1$ , where  $E[G(\chi)] \equiv \int d\chi G(\chi) \exp(-S[\chi]) / \int d\chi \exp(-S[\chi])$ .  
[6] A. Zee, "Quantum field theory in a nutshell," (Princeton University Press, 2010) Chap. VI, 2nd ed.  
[7] C. Wetterich, Physics Letters B **301**, 90 (1993).

[8] We use the following matrix-form notation to denote the spatio-temporal (or momentum space) integration of fields:

$$\begin{aligned} \int d^d x \sum_n J_n(x) \chi_n(x) &\rightarrow J^T \chi, \\ \frac{1}{2} \int \frac{d^d q}{(2\pi)^d} \sum_{m,n} \chi_m(q) (R_k)_{m,n}(q) \chi_n(-q) &\rightarrow \frac{1}{2} \chi^T R_k \chi, \\ \frac{\delta S}{\delta \phi(x)}[\phi] &\rightarrow \frac{\delta S}{\delta \phi}[\phi], \end{aligned}$$

where  $d$  is the dimension of the space-time, and  $m, n$  are the indices of the components.

- [9] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, The Journal of Chemical Physics **21**, 1087 (1953).
- [10] W. K. Hastings, Biometrika **57**, 97 (1970).
- [11] G. O. Roberts and R. L. Tweedie, Bernoulli, 341 (1996).
- [12] In particular, strong-constraint 4D-Var has

$$F[\phi] = \begin{bmatrix} B^{-\frac{1}{2}} (\phi - m_0) \\ R^{-\frac{1}{2}} (y - \Psi[\phi]) \end{bmatrix},$$

where  $\Psi$  is the model,  $B, R$  are the background and observational error covariance matrices, respectively, and  $m_0, y$  are the first-guess field and observational data, respectively.

- [13] R. M. May *et al.*, Nature **261**, 459 (1976).
- [14] K. Law, A. Stuart, and K. Zygalakis, *Data Assimilation: A Mathematical Introduction* (Springer, 2015).
- [15] In this case,  $F[\phi] = [(\phi - m_0)/\sigma_0, (y_1 - \Psi^{(1)}[\phi])/\gamma, \dots, (y_J - \Psi^{(J)}[\phi])/\gamma]^T$ .
- [16] E. Lorenz, in *Seminar on Predictability, 4-8 September 1995*, Vol. 1, ECMWF (ECMWF, Shinfield Park, Reading, 1995) pp. 1–18.
- [17] A. C. Lorenc and T. Payne, Quarterly Journal of the Royal Meteorological Society **133**, 607 (2007).